The capacity of hybrid quantum memory

Greg Kuperberg[1,*]
[1]UC Davis

The general stable quantum memory unit is a hybrid consisting of a classical digit with a quantum digit (qudit) assigned to each classical state. The shape of the memory is the vector of sizes of these qudits, which may differ. We determine when N copies of a quantum memory $\mathscr{A}$ embed in $N(1+o(1))$ copies of another quantum memory $\mathscr{B}$. This relationship captures the notion that $\mathscr{B}$ is as at least as useful as $\mathscr{A}$ for all purposes in the bulk limit. The answer is that the embeddings exist if and only if for all p, the p-norm of the shape of $\mathscr{A}$ does not exceed the p-norm of the shape of $\mathscr{B}$. The log of the p-norm of the shape of $\mathscr{A}$ can be interpreted as the maximum of $S(\rho)+H(\rho)/p$ (quantum entropy plus discounted classical entropy) taken over all mixed states ρ on $\mathscr{A}$. Thus the bulk utility of a quantum memory is determined by its simultaneous capacity for classical and quantum entropy, which is not a finite list of numbers, but rather a convex region in the classical-quantum entropy plane.

1. INTRODUCTION

Many questions in quantum information theory involve both quantum and classical information. The usual computational model for such dual information is independent quantum and classical memory. The measurement algebra of a combined memory consisting of an a-state qudit and a b-state classical digit is

$$\mathscr{M}_a \otimes \mathbb{C}^b = \bigoplus_{k=1}^{b} \mathscr{M}_a,$$

where $\mathscr{M}_a$ is the set of $a \times a$ matrices. But this is not the most general possible hybrid of classical and quantum memory. Rather the measurement algebra $\mathscr{A}$ of a finite memory could be any direct sum of matrix algebras of possibly different dimensions:

$$\mathscr{A} \cong \bigoplus_{k=1}^{n} \mathscr{M}_{\lambda_k}.$$

The partition (*i.e.*, non-negative integral vector) $\lambda = \lambda(\mathscr{A})$ is a list of the dimensions of the matrix algebras called the *shape* of the memory $\mathscr{A}$. (Section 2 discusses why this is a reasonably general quantum memory model.)

For example, the simplest hybrid memory is a *hybrid trit*, with shape $(2,1)$. It consists of matrices of the form

$$\left(\begin{array}{cc|c} * & * & 0 \\ * & * & 0 \\ \hline 0 & 0 & * \end{array}\right).$$

This memory models a three-state system which one state is observed by the environment but the other two remain coherent relative to each other. It is easy to compare the capacity of the hybrid trit to any other quantum memory: It is between a qubit and a qutrit, more than a classical trit, less than any larger memory that contains a qubit, and neither more nor less than a classical digit with at least 4 states.

It turns out that there is more than one notion by which one memory unit has more capacity than another. (Atypically, all such notions are equivalent for the hybrid trit.) The strictest relevant relationship between memories is given by algebra embeddings. A memory $\mathscr{B}$ can hold as much state as another memory $\mathscr{A}$ if and only if there is an algebra embedding $\mathscr{A} \hookrightarrow \mathscr{B}$. The embedding need not be unit-preserving (unital). As Section 3.1 explains, the question of whether $\mathscr{A}$ embeds in $\mathscr{B}$ is a computable (but NP-hard) bin-packing problem.

In this article we will consider a more relaxed comparison, namely whether many copies of $\mathscr{A}$ embed in slightly more copies of $\mathscr{B}$. More precisely we say that $\mathscr{A}$ *bulk-embeds* in $\mathscr{B}$, or $\mathscr{A} \overset{b}{\hookrightarrow} \mathscr{B}$, if for every rational $\varepsilon > 0$, there exists an N such that

$$\mathscr{A}^{\otimes N} \hookrightarrow \mathscr{B}^{\otimes N(1+\varepsilon)}.$$

If $\mathscr{A}$ bulk-embeds in $\mathscr{B}$, there is no reason to pay more for $\mathscr{A}$ than $\mathscr{B}$ when buying large quantities of the two memories with equal performance. Our main result is a characterization of when $\mathscr{A}$ bulk-embeds in $\mathscr{B}$:

Theorem 1. *If $\mathscr{A}$ and $\mathscr{B}$ are two hybrid memories, then $\mathscr{A} \overset{b}{\hookrightarrow} \mathscr{B}$ if and only if*

$$||\lambda(\mathscr{A})||_p \leq ||\lambda(\mathscr{B})||_p$$

for all $p \in [1,\infty]$.

One direction of Theorem 1 is straightforward. The p-norm of a partition λ is defined as

$$||\lambda||_p = \left(\sum_k \lambda_k^p\right)^{1/p}.$$

It is easy to check that the p-norm is multiplicative:

$$||\lambda(\mathscr{A} \otimes \mathscr{B})||_p = ||\lambda(\mathscr{A})||_p ||\lambda(\mathscr{B})||_p$$

for any pair of memories $\mathscr{A}$ and $\mathscr{B}$. On the other hand the bin-packing model implies that if $\mathscr{A}$ embeds in $\mathscr{B}$, then

$$||\lambda(\mathscr{A})||_p \leq ||\lambda(\mathscr{B})||_p.$$

*Electronic address: greg@math.ucdavis.edu; Supported by NSF grant DMS #0072342

It follows that this inequality also holds when $\mathscr{A}$ bulk-embeds in $\mathscr{B}$. The proof of the other direction of Theorem 1 is the topic of Section 3.

The p-norm has an interesting information-theoretic interpretation. In Section 4 we will define the classical entropy $H(\rho)$ and the quantum entropy $S(\rho)$ of a state ρ of a quantum memory $\mathscr{A}$. Their definitions are justified by an encoding theorem due to Barnum, Hayden, Jozsa, and Winter [1], and by a capacity estimate:

Theorem 2. *Every state ρ of a memory $\mathscr{A}$ satisfies inequality*

$$\frac{H(\rho)}{p} + S(\rho) \leq \log ||\lambda(\mathscr{A})||_p,$$

where ρ has classical entropy $H(\rho)$ and quantum entropy $S(\rho)$. For each p there exists a ρ that achieves equality. Any non-negative pair (H,S) satisfying the inequality for all p can be expressed as

$$(H,S) = (H(\rho) + tS(\rho), (1-t)S(\rho))$$

for some ρ and some $t \in [0,1]$.

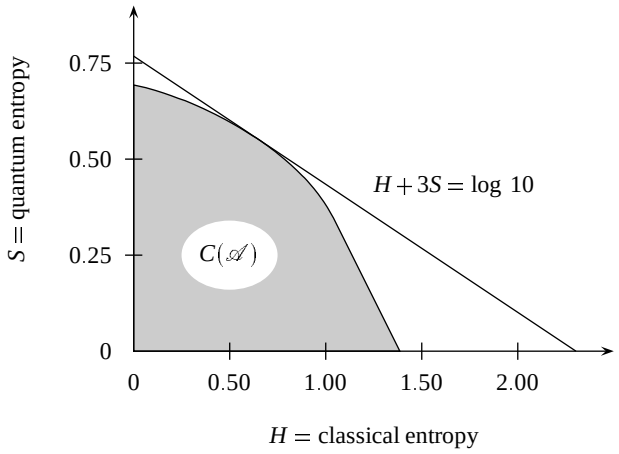

Figure 1: The capacity region of a memory $\mathscr{A}$ with shape $(2,1,1)$, and its 3-norm bounding line.

Thus the set of possible pairs

$$(H(\rho) + tS(\rho), (1-t)S(\rho))$$

for states ρ on a memory $\mathscr{A}$ and constants $t \in [0,1]$ forms a convex capacity region $C(\mathscr{A})$ in the first quadrant of the plane. Figure 1 shows an example. The presence of the constant t expresses the fact that quantum entropy can be used to communicate classical information. The S-intercept of the line tangent to $C(\mathscr{A})$ with slope $-\frac{1}{p}$ is $\log ||\lambda(\mathscr{A})||_p$. A memory $\mathscr{A}$ bulk-embeds in another memory $\mathscr{B}$ if and only if $C(\mathscr{A}) \subseteq C(\mathscr{B})$. In other words, $\mathscr{A}$ bulk-embeds in $\mathscr{B}$ if and only if it has no state ρ with too much entropy to fit in $\mathscr{B}$.

Note that the three values of p that are generally important for p-norms each have a special significance for hybrid memories. The logarithm of the 1-norm, $\log ||\lambda(\mathscr{A})||_1$, is the purely classical capacity of $\mathscr{A}$. The logarithm of the ∞-norm, $\log ||\lambda(\mathscr{A})||_\infty$, is the purely quantum capacity. Finally the logarithm of the 2-norm,

$$\log ||\lambda(\mathscr{A})||_2 = \frac{\log \dim \mathscr{A}}{2},$$

is half of the dense coding capacity of $\mathscr{A}$.

2. MEMORY

As suggested in the introduction, the first question is whether the proposed model of a hybrid memory is adequately general. One justification comes from viewing a quantum system not as a Hilbert space, but as an abstract operator algebra $\mathscr{A}$. If $\mathscr{A}$ is infinite-dimensional, it should satisfy some analytic axioms in order to be useful for quantum probability theory; usually it is assumed to be either a C^*-algebra or a von Neumann algebra [6, 7]. But if it is finite-dimensional, it suffices to require that $\mathscr{A}$ be a (positive-definite) $*$-*algebra*; it is then also a C^*-algebra and a von Neumann algebra. This means that in addition to the fact that $\mathscr{A}$ is a complex vector space with associative multiplication, it has an abstract $*$-operation which is anti-linear, product-reversing, and suitably positive-definite:

$$(\lambda AB)^* = \overline{\lambda} B^* A^* \qquad A^*A = 0 \implies A = 0.$$

For example, any direct sum of matrix algebras is a $*$-algebra.

Despite their abstraction, $*$-algebras have all of the necessary structure for quantum information theory. The elements of a $*$-algebra $\mathscr{A}$ of the form A^*A are called *positive*. A *state* ρ on a $*$-algebra $\mathscr{A}$ is defined as a dual vector $\rho \in \mathscr{A}^*$ which is positive on positive elements and which is normalized by $\rho(I) = 1$. Consequently we write $\rho(A)$ for the expectation of A rather than $\mathrm{Tr}(\rho A)$. A *quantum operation* from a system with $*$-algebra $\mathscr{A}$ to a system with $*$-algebra $\mathscr{B}$ is defined as a completely positive, unital (CPU) linear map $\mathscr{E} : \mathscr{A} \to \mathscr{B}$. Here *completely positive* means that $\mathscr{E}$ sends positive elements to positive elements after tensoring with the identity on a third $*$-algebra. Note that the transpose $\mathscr{E}^T : \mathscr{B}^* \to \mathscr{A}^*$ is the corresponding map on states. It is completely positive and trace-preserving if we take $\rho(I)$ to be the trace of ρ.

A standard classification theorem [2] says that every finite-dimensional $*$-algebra $\mathscr{A}$ is a direct sum of matrix algebras,

$$\mathscr{A} \cong \bigoplus_{k=1}^{n} \mathscr{M}_{\lambda_k}.$$

Thus a quantum memory of shape λ is the most general possible finite-dimensional complex algebra of observables satisfying reasonable algebraic axioms. (However abandoning $\mathbb{C}$ as the field of scalars leads to other possibilities [3].)

Another justification comes from the interaction of a physical memory with its environment. Consider a physical device whose state is defined by a $*$-algebra $\mathscr{M}$. Realistically $\mathscr{M}$ is very large, but almost all of it is thermally coupled to the environment. Its decoherence on the thermal time scale is given

by some quantum operation $\mathscr{E} : \mathscr{M} \to \mathscr{M}$. If the thermal time scale is much shorter than the computational time scale, then the information retained by $\mathscr{E}^n$ in the limit $n \to \infty$ is the reliable memory of $\mathscr{M}$.

Certainly any finite-dimensional $*$-algebra $\mathscr{A}$ is the reliable memory retained by some quantum operation on a matrix algebra $\mathscr{M}_d$. In the minimal construction, let $d = ||\lambda(\mathscr{A})||_1$ be the total size of all blocks of $\mathscr{A}$ and realize $\mathscr{A} \subseteq \mathscr{M}_d$ as matrices that consist of diagonal blocks of size $\lambda_k(\mathscr{A})$ for each k. Then there is a POVM whose kth element P_k is the identity of the kth summand $\mathscr{A}_k$. The corresponding quantum operation

$$\mathscr{P}(A) = \sum_{k=1}^{n} P_k A P_k$$

is a projection, meaning $\mathscr{P}^2 = \mathscr{P}$, and its image is $\mathscr{A}$. If the thermal evolution of $\mathscr{M}_d$ is given by $\mathscr{P}$, the algebra $\mathscr{A}$ defines the retained information.

Conversely, the following two results establish that, for any quantum operation $\mathscr{E}$ on a finite-dimensional $*$-algebra, the information retained by $\mathscr{E}^n$ in the limit $n \to \infty$ is defined by a smaller $*$-algebra of effective observables. (See also Zurek [10].)

Theorem 3. *Let $\mathscr{E} : \mathscr{M} \to \mathscr{M}$ be a CPU map on a finite-dimensional $*$-algebra $\mathscr{M}$. Then there exists a sequence of integers $n_k \to \infty$ such that $\mathscr{E}^{n_k}$ converges to a unique projection $\mathscr{P}$.*

Proof. (Sketch) Choose a basis of $\mathscr{M}$ that puts $\mathscr{E}$ in Jordan canonical form. Since $\mathscr{E}^n$ is CPU, its matrix entries are bounded. Therefore $\mathscr{E}$ has no eigenvalues λ with $|\lambda| > 1$, and if $|\lambda| = 1$, the λ-isotypic part of $\mathscr{E}$ is diagonal. Choose a sequence of exponents $n_k \to \infty$ such that the phases of these diagonal entries of $\mathscr{E}^{n_k}$ are aligned with 1 in the limit. The rest of the matrix of $\mathscr{E}^n$ decays to 0 as $n \to \infty$. The map $\mathscr{P}$ is unique because if the phases do not align with 1, the limiting map is not a projection. □

Theorem 4 (Choi, Effros [4, p.166-7]). *If $\mathscr{M}$ is a finite-dimensional $*$-algebra and $\mathscr{P}$ is a CPU projection on $\mathscr{M}$, then the image of $\mathscr{P}$ is a $*$-algebra $\mathscr{A}$ with a modified product $A \circ B = \mathscr{P}(AB)$.*

The non-trivial part of Theorem 4 is the fact that the modified product $A \circ B$ is associative. It is generally not associative if $\mathscr{P}$ is not simultaneously completely positive, unital, and a projection. (However it is enough to require that $\mathscr{P}$ be 2-positive.) The modified product structure is consistent with applying $\mathscr{P}$ between any two computational manipulations of $\mathscr{M}$.

3. EMBEDDINGS

3.1. Bin packing

Besides embeddability and bulk embeddability, we will also compare memories using a partial ordering on partitions which resembles dominance [9, Ch.7], or majorization, but which is actually stricter. The partition λ *supermajorizes* the partition μ, or $\mu \preccurlyeq_S \lambda$, means that for every n, the sum of all parts of λ that are at least n exceeds the same sum for μ. Supermajorization lies between embeddability and bulk embeddability:

$$\mathscr{A} \hookrightarrow \mathscr{B} \implies \lambda(\mathscr{A}) \preccurlyeq_S \lambda(\mathscr{B}) \implies \mathscr{A} \overset{b}{\hookrightarrow} \mathscr{B}$$
$$\mathscr{A} \overset{b}{\hookrightarrow} \mathscr{B} \not\Longrightarrow \lambda(\mathscr{A}) \preccurlyeq_S \lambda(\mathscr{B}) \not\Longrightarrow \mathscr{A} \hookrightarrow \mathscr{B}.$$

We can view the parts of a partition λ as an unordered multiset $\{\lambda_k\}$. It is sometimes convenient to assume a specific order on the parts. In this case we follow the usual convention that the parts of λ are non-increasing:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 1.$$

Given a partition λ, let λ_x denote the sum of all parts of λ that are at least x. Thus the relationship $\lambda \preccurlyeq_S \mu$ means that

$$\lambda_{\geq x} \leq \mu_{\geq x}$$

for all x. Obviously it suffices to consider integer values of x, but it will be cleaner later to allow non-integer values. Also $\ell\lambda$ denotes λ with each part repeated ℓ times. (This is not to be confused with magnifying each part by a factor of ℓ.)

In order to analyze bulk embeddings and prove Theorem 1, we first analyze ordinary embeddings [2]. If $\mathscr{A}$ and $\mathscr{B}$ are finite-dimensional $*$-algebras, then any algebra homomorphism $f : \mathscr{A} \to \mathscr{B}$ is characterized by a *Bratteli diagram* Γ whose vertices are the summands of $\mathscr{A}$ and $\mathscr{B}$. Let $\mathscr{A}_k$ be the kth summand of $\mathscr{A}$, so that $\mathscr{A}_k \cong \mathscr{M}_{\lambda_k}$, and likewise for $\mathscr{B}$. If we denote the adjacency matrix of Γ by Γ as well, then the diagram's interpretation is that f embeds $\Gamma_{j,k}$ copies of $\mathscr{A}_j$ in $\mathscr{B}_k$. (The matrix Γ is the adjacency matrix of the diagram Γ.) The matrix Γ must satisfy the inequality

$$\sum_j \Gamma_{j,k} \lambda(\mathscr{A})_j \leq \lambda(\mathscr{B})_k$$

for all k. (Bratteli diagrams often describe unital homomorphisms, which require equality.) The homomorphism f is an embedding if and only if each summand of $\mathscr{A}$ has at least one edge, or equivalently that

$$\sum_k \Gamma_{j,k} \geq 1$$

for all j.

Thus we can think of $\mathscr{A}$ as a set of 1-dimensional blocks, $\mathscr{B}$ as a set of 1-dimensional bins, and the embedding as a way to pack the blocks of $\mathscr{A}$ in the bins of $\mathscr{B}$. The embedding might replicate some of the summands of $\mathscr{A}$ before packing them, but if there is any embedding, there is one that does not replicate any summands. (Replication in this sense has nothing to do with cloning as in the no-cloning theorem.)

Lemma 5. *If $\mathscr{A} \hookrightarrow \mathscr{B}$, then $\lambda(\mathscr{A}) \preccurlyeq_S \lambda(\mathscr{B})$. If $2\lambda(\mathscr{A}) \preccurlyeq_S \mu(\mathscr{B})$, then $\mathscr{A} \hookrightarrow \mathscr{B}$.*

Proof. Both statements follow by induction on the number of parts of $\lambda(\mathscr{A})$. They both hold trivially when $\lambda(\mathscr{A})$ is empty. To prove the first assertion, suppose that in some embedding, $\mathscr{A}_1$ embeds in $\mathscr{B}_k$. Let $\widehat{\mathscr{A}}$ be $\mathscr{A}$ with $\mathscr{A}_1$ removed and let $\widehat{\mathscr{B}}$ be $\mathscr{B}$ with $\mathscr{B}_k$ reduced by $\lambda(\mathscr{A})_1$, or removed if $\lambda(\mathscr{B})_k = \lambda(\mathscr{A})_1$. Then by construction, $\widehat{\mathscr{A}} \hookrightarrow \widehat{\mathscr{B}}$. Thus by induction,

$$\lambda(\widehat{\mathscr{A}})_{\geq x} \leq \lambda(\widehat{\mathscr{B}})_{\geq x}$$

for all $x \geq 1$. By the definition λ' and μ',

$$\lambda(\widehat{\mathscr{A}})_{\geq x} = \lambda(\mathscr{A})_{\geq x} - \lambda(\mathscr{A})_1$$
$$\lambda(\widehat{\mathscr{B}})_{\geq x} \leq \lambda(\mathscr{B})_{\geq x} - \lambda(\mathscr{A})_1$$

for $x \leq \lambda(\mathscr{A})_1$, while $\lambda(\mathscr{A})_{\geq x}$ vanishes for $x > \lambda(\mathscr{A})_1$. Thus

$$\lambda(\mathscr{A})_{\geq x} \leq \lambda(\mathscr{B})_{\geq x},$$

as desired.

To prove the second assertion, suppose that $2\lambda(\mathscr{A}) \preccurlyeq_S \mu$, or equivalently that

$$2\lambda(\mathscr{A})_{\geq x} \leq \lambda(\mathscr{B})_{\geq x}$$

for all x. We can greedily put $\mathscr{A}_1$ in any $\mathscr{B}_k$ in which it fits and make $\widehat{\mathscr{A}}$ and $\widehat{\mathscr{B}}$ as before. (Note that in this greedy algorithm it is important to start with the largest summand of $\mathscr{A}$, not an arbitrary one.) If $\lambda(\mathscr{B})_k \leq 2\lambda(\mathscr{A})_1$, then

$$\lambda(\widehat{\mathscr{A}})_{\geq x} = \lambda(\mathscr{A})_{\geq x} - \lambda(\mathscr{A})_1$$
$$\lambda(\widehat{\mathscr{B}})_{\geq x} \geq \lambda(\mathscr{B})_{\geq x} - 2\lambda(\mathscr{A})_1$$

for all $x \leq \lambda(\mathscr{A})_1$, while $\lambda(\widehat{\mathscr{A}})_{\geq x}$ vanishes for $x > \lambda(\mathscr{A})_1$. On the other hand if $\lambda(\mathscr{B})_k \geq 2\lambda(\mathscr{A})_1$, then bin k remains larger than any block even after block 1 is subtracted. In this case

$$\lambda(\widehat{\mathscr{A}})_{\geq x} = \lambda(\mathscr{A})_{\geq x} - \lambda(\mathscr{A})_1$$
$$\lambda(\widehat{\mathscr{B}})_{\geq x} \geq \lambda(\mathscr{B})_{\geq x} - \lambda(\mathscr{A})_1$$

for all $x \leq \lambda(\mathscr{A})_1$. Thus

$$2\lambda(\widehat{\mathscr{A}}) \preccurlyeq_S \lambda(\widehat{\mathscr{B}})$$

either way, so the bin packing exists by induction. □

3.2. Large deviations

The proof of Theorem 1 combines Lemma 5 with the Chernoff-Cramér theorem on large deviations [5]. The theorem is usually stated in terms of sums of independent random variables, but it is more convenient here to formulate it in terms of convolutions of measures:

Theorem 6 (Chernoff, Cramér). *Let μ be a measure on an interval $[0,u]$, let*

$$\ell(\beta) = \log \int_0^\infty e^{\beta x} \mu$$

be the logarithm of the Laplace transform of μ, let $t > 0$, Then for all $n \in \mathbb{Z}_+$ and all $\beta > 0$,

$$\int_{nt}^\infty \mu^{*n} \leq e^{n(\ell(\beta) - \beta t)}$$

If $\ell'(0) \leq t < u$ and β minimizes

$$\ell(\beta) - \beta t,$$

then for all $0 < s < t$,

$$\int_{n(t-s)}^\infty \mu^{*n} \geq e^{n(\ell(\beta) - \beta t - \beta s)} \left(1 - \frac{\ell''(\beta)}{ns^2}\right).$$

Here μ^{*n} denotes the n-fold convolution of μ with itself. When $\ell'(0) < t < u$, the expression

$$\widehat{\ell}(t) = \min_\beta \ell(\beta) - \beta t$$

is the Legendre transform of $\ell(\beta)$. Note that a unique β achieves the minimum because the minimand is concave up, increases as $\beta \to \infty$, and does not increase at $\beta = 0$.

Sketch. For any β,

$$\begin{aligned}\int_{nt}^\infty \mu^{*n} &\leq e^{-n\beta t} \int_0^\infty e^{\beta x} \mu^{*n} \\ &= e^{-n\beta t} e^{n\ell(\beta)}.\end{aligned}$$

This establishes the upper bound, Chernoff's inequality.

If β is chosen to minimize $\ell(\beta) - \beta t$, then $t = \ell'(\beta)$. In this case

$$\begin{aligned}\int_{n(t-s)}^\infty \mu^{*n} &\geq e^{-n\beta(s+t)} \int_{n(t-s)}^{n(t+s)} e^{\beta x} \mu^{*n} \\ &\geq e^{-n\beta(s+t)} \int_0^\infty \left(1 - \frac{(x-nt)^2}{(ns)^2}\right) e^{\beta x} \mu^{*n} \\ &= e^{-n\beta(s+t)} \left(1 - \frac{\ell''(\beta)}{ns^2}\right) e^{n\ell(\beta)}.\end{aligned}$$

The equality uses the identities

$$\int_0^\infty x e^{\beta x} \mu^{*n} = \left(e^{n\ell(\beta)}\right)' = n\ell'(\beta) e^{n\ell(\beta)}$$
$$\int_0^\infty x^2 e^{\beta x} \mu^{*n} = \left(e^{n\ell(\beta)}\right)'' = \left(n\ell''(\beta) + n^2 \ell'(\beta)^2\right) e^{n\ell(\beta)}.$$

This establishes the lower bound, Cramér's theorem. □

Proof of Theorem 1. In brief, without loss of generality

$$||\lambda(\mathscr{A})||_p < ||\lambda(\mathscr{B})||_p$$

for all $p \in [1,\infty]$. In this case we apply Theorem 6 to the measures

$$\mu_{\mathscr{A}} = \sum_k \lambda_k \delta_{\log \lambda_k(\mathscr{A})}$$
$$\mu_{\mathscr{B}} = \sum_k \lambda_k \delta_{\log \lambda_k(\mathscr{B})},$$

where δ_x denotes a delta function (or atom) at x. For sufficiently large n, Chernoff's bound for $\mu_{\mathscr{A}}$ and Cramér's inequality for $\mu_{\mathscr{B}}$ together imply the criterion

$$2\lambda(\mathscr{A}^{\otimes n})_{\geq x} \leq \lambda(\mathscr{B}^{\otimes n})_{\geq x}$$

of Lemma 5 uniformly for $x \in [1,\infty)$.

In detail, we assume that $||\lambda(\mathscr{B})||_\infty > 1$; otherwise $\mathscr{A}$ and $\mathscr{B}$ are both entirely classical and Theorem 1 is easy. Since

$$||\lambda(\mathscr{A})||_p \leq ||\lambda(\mathscr{B})||_p$$

for all $p \in [1,\infty]$, then for any $k > 1$,

$$||\lambda(\mathscr{A}^{\otimes k})||_p < ||\lambda(\mathscr{B}^{\otimes k+1})||_p.$$

The ε margin in Theorem 1 thus allows us to assume that

$$||\lambda(\mathscr{A})||_p < ||\lambda(\mathscr{B})||_p$$

for all $p \in [1,\infty]$ by replacing $\mathscr{A}$ by $\mathscr{A}^{\otimes k}$ and $\mathscr{B}$ by $\mathscr{B}^{\otimes k+1}$.

The measure $\mu_{\mathscr{A}}$ is defined so that

$$\mu_{\mathscr{A}}^{*n} = \lambda(\mathscr{A}^{\otimes n})$$

and

$$\lambda(\mathscr{A})_{\geq x} = \int_{e^x}^\infty \mu_{\mathscr{A}},$$

and likewise for $\mu_{\mathscr{B}}$. Therefore by Lemma 5, it suffices to show that there exists an n such that for all $t \geq 0$,

$$2\int_{nt}^\infty \mu_{\mathscr{A}} \leq \int_{nt}^\infty \mu_{\mathscr{B}}. \tag{1}$$

As in the statement of Theorem 6, let

$$\ell_{\mathscr{A}}(\beta) = \log \int_0^\infty e^{\beta x}\mu_{\mathscr{A}} = \log||\lambda(\mathscr{A})||_{\beta+1}$$
$$\ell_{\mathscr{B}}(\beta) = \log \int_0^\infty e^{\beta x}\mu_{\mathscr{B}} = \log||\lambda(\mathscr{B})||_{\beta+1}.$$

Observe that $\ell_{\mathscr{B}}(\beta)$ is a smooth, concave function, and that

$$\lim_{\beta\to\infty} \frac{\ell'_{\mathscr{B}}(\beta)}{\beta} = \log||\lambda(\mathscr{B})||_\infty < \infty.$$

It follows that $\ell''_{\mathscr{B}}(\beta)$ has a finite maximum C for $\beta \in [0,\infty)$. Note also that

$$\frac{\ell_{\mathscr{B}}(\beta) - \ell_{\mathscr{A}}(\beta)}{\beta}$$

achieves a positive minimum, since

$$\lim_{\beta\to\infty} \frac{\ell_{\mathscr{B}}(\beta) - \ell_{\mathscr{A}}(\beta)}{\beta} = ||\lambda(\mathscr{B})||_\infty - ||\lambda(\mathscr{A})||_\infty$$
$$\lim_{\beta\to 0} \frac{\ell_{\mathscr{B}}(\beta) - \ell_{\mathscr{A}}(\beta)}{\beta} = \infty.$$

Temporarily suppose that $t \geq \ell'_{\mathscr{B}}(0)$ and that $\beta = \beta(t)$ minimizes $\ell_{\mathscr{B}}(\beta) - \beta t$. Let

$$s = \sqrt{\frac{2C}{n}}.$$

Then

$$\int_{n(t-s)}^\infty \mu_{\mathscr{A}}^{*n} \leq e^{n(\ell_{\mathscr{A}}(\beta)-\beta t+\beta s)}$$
$$\int_{n(t-s)}^\infty \mu_{\mathscr{B}}^{*n} \geq e^{n(\ell_{\mathscr{B}}(\beta)-\beta t-\beta s)-\log 2}.$$

If n is large enough that

$$2s + \frac{2\log 2}{n} = 2\sqrt{\frac{2C}{n}} + \frac{2\log 2}{n} \leq \min_\beta \frac{\ell_{\mathscr{B}}(\beta) - \ell_{\mathscr{A}}(\beta)}{\beta},$$

then

$$2\int_{n(t-s)}^\infty \mu_{\mathscr{A}}^{*n} \leq \int_{n(t-s)}^\infty \mu_{\mathscr{B}}^{*n}.$$

Thus for some $\varepsilon > 0$, inequality (1) holds for all $t > \ell'_{\mathscr{B}}(0) - \varepsilon$.

If $t \leq \ell'_{\mathscr{B}}(0) - \varepsilon$, let $u = \ell'_{\mathscr{B}}(0)$ and let $\beta = 0$. Then

$$\int_{nt}^\infty \mu_{\mathscr{A}}^{*n} \leq \int_0^\infty \mu_{\mathscr{A}}^{*n} = e^{n\ell_{\mathscr{A}}(0)},$$

while

$$\int_{nt}^\infty \mu_{\mathscr{B}}^{*n} \geq \int_{n(u-s)}^\infty \mu_{\mathscr{B}}^{*n} \geq e^{n\ell_{\mathscr{B}}(0)-\log 2}$$

provided that $s \leq \varepsilon$. Since $\ell_{\mathscr{A}}(0) < \ell_{\mathscr{B}}(0)$, inequality (1) holds when n is large enough. □

4. ENTROPY

Let $\mathscr{A}$ be a finite-dimensional $*$-algebra, where as before

$$\mathscr{A} = \bigoplus_{k=1}^n \mathscr{A}_k \cong \bigoplus_{k=1}^n \mathscr{M}_{\lambda_k}.$$

Let ρ be a (mixed) state on $\mathscr{A}$; as explained above we view ρ as a dual vector on $\mathscr{A}$ rather than as an element. Let

$$\rho_k = \rho|_{\mathscr{A}_k}$$

be the restriction of ρ to $\mathscr{A}_k$. Diagonalize each ρ_k and let $r_{k,j}$ with $1 \leq j \leq \lambda_k$ be its diagonal entries. Let

$$r_k = \rho_k(I) = \sum_{j=1}^{\lambda_k} r_{k,j}$$

be the total density of ρ in $\mathscr{A}_k$; evidently

$$\sum_{k=1}^n r_k = 1.$$

It is also convenient to define the normalized state ρ'_k on $\mathscr{A}_k$ by

$$\rho'_k = \frac{\rho_k}{r_k},$$

with diagonal entries

$$r'_{k,j} = \frac{r_{k,j}}{r_k}.$$

The *classical entropy* of the state ρ is defined as

$$H(\rho) = -\sum_{k=1}^{n} r_k \log\, r_k.$$

The *quantum entropy* of ρ is defined as

$$S(\rho) = -\sum_{k=1}^{n}\sum_{j=1}^{\lambda_k} r_{k,j} \log\, r'_{k,j}.$$

These two entropies are supported by a number of elementary justifications: The classical entropy of ρ is the Shannon entropy of the restriction of ρ to the center of $\mathscr{A}$, which is a classical system. The quantum entropy of ρ is the expected value of the von Neumann entropy of ρ_k, where the index k is chosen randomly with probability r_k. Finally the total entropy

$$H(\rho) + S(\rho) = -\sum_{k=1}^{n}\sum_{j=1}^{\lambda_k} r_{k,j} \log\, r_{k,j}$$

has the same formula as both the Shannon and the von Neumann entropy. These definitions of entropy can also be justified by a rate theorem given at the end of the section.

The proof of Theorem 2 is based on finding thermal states of $\mathscr{A}$ with respect to a certain Hamiltonian. We define the energy E_k of the summand $\mathscr{A}_k$ as the negative of its capacity for quantum entropy:

$$E_k = -\log \lambda_k(\mathscr{A}).$$

We retain the parameter β from Section 3.2, setting $p = \beta + 1$, and we also define the temperature $T = 1/\beta$. The thermal state ρ_T at temperature T has the property that its restriction ρ_k to each $\mathscr{A}_k$ is uniform. If ρ is any state with this property, then its energy $E(\rho)$ is, by definition, the negative of its quantum entropy:

$$E(\rho) = -S(\rho).$$

The free energy of ρ is therefore

$$F(\rho) = E(\rho) - T(H(\rho) + S(\rho)) = -T(H(\rho) + pS(\rho)).$$

Since the thermal state minimizes the free energy, we have defined energy so that the thermal state $\rho(T)$ maximizes quantum entropy plus classical entropy discounted by p. To compute the maximum recall that for the thermal state $\rho(T)$, the free energy is proportional to the log of the partition function:

$$F(\rho(T)) = -T \log\, Z(\rho(T)) = -T \log \left(\sum_{k=1}^{n} \lambda_k e^{\beta \log\, \lambda_k(\mathscr{A})} \right)$$
$$= -T \log \sum_{k=1}^{n} \lambda_k^{\beta+1} = -T\, p \log\, ||\lambda(\mathscr{A})||_p.$$

Therefore

$$\frac{H(\rho(T))}{p} + S(\rho(T)) = \log\, ||\lambda(\mathscr{A})||_p,$$

as desired.

Theorem 7. *Let $\mathscr{A}$ be a quantum memory with a state ρ and let $\mathscr{B}$ be another quantum memory. Then there are reliable compression schemes*

$$\mathscr{A}^{\otimes N} \xrightarrow{\mathscr{X}_N} \mathscr{B}^{\otimes N(1+\varepsilon)} \xrightarrow{\mathscr{Y}_N} \mathscr{A}^{\otimes N}$$

for every rational $\varepsilon > 0$ if and only if $(H(\rho), S(\rho)) \in C(\mathscr{B})$.

Proof. (Sketch) Barnum, Hayden, Jozsa, and Winter establish the special case in which $\mathscr{B}$ consists of a qudit and a separate classical digit [1]. In this case $\lambda(\mathscr{B})$ is a rectangle, while the capacity region $C(\mathscr{B})$ is a trapezoid with an obtuse corner $(H(\mathscr{B}), S(\mathscr{B}))$. If $\mathscr{C}$ is a more general target memory such that $(H(\mathscr{B}), S(\mathscr{B})) \in C(\mathscr{C})$, then $\mathscr{B}$ bulk-embeds in $\mathscr{C}$. Since bulk embedding is itself a form of reliable encoding, composing the two encodings establishes the "go" direction of the theorem.

The "no-go" direction follows similarly. Suppose that $\mathscr{B}$ is a general target memory such that $(H(\rho), S(\rho)) \notin C(\mathscr{B})$ but the state $\rho^{\otimes N}$ can be compressed reliably into $\mathscr{B}^{\otimes N(1+\varepsilon)}$. Then its image under compression has a monographic state σ such that $(H(\sigma), S(\sigma)) \in C(\mathscr{B})$. But then $\sigma^{\otimes N(1+\varepsilon)}$ can be compressed into a rectangular memory which is too small to hold $\rho^{\otimes N}$. □

5. DISCUSSION

Our results motivate some philosophical conclusions and open problems.

Section 2 illustrates the principle that classical information theory is the abelian special case of quantum information theory. Some authors maintain a dichotomy between the two theories by considering ensembles of mixed states. But such notation is ultimately redundant, because an ensemble is itself a classical probabilistic state. More precisely, let

$$\rho = \sum_k p_k \rho_k \in \mathscr{A}$$

be an ensemble of states in a memory $\mathscr{A}$. If the symbol k is not recorded, then, as is well-known, ρ encodes all statistical information that can be extracted from the ensemble. But if

each symbol k is recorded as a state σ_k in another memory $\mathscr{B}$, then we can let

$$\rho' = \sum_k p_k \rho_k \otimes \sigma_k \in \mathscr{A} \otimes \mathscr{B}.$$

If $\mathscr{B}$ is abelian and the σ_k's are distinct pure states, then the state ρ' is equivalent to an ensemble with a record of its preparation.

Theorems 1 and 2 together suggest that all quantum information can be measured in the bulk limit by two numbers, classical entropy H and quantum entropy S. But information capacity has more structure than information itself. The capacity of a quantum memory is defined by a curve that represents trade-offs between classical and quantum entropy. The capacity of a general quantum channel could be even more complicated.

There are many interesting partial orderings on quantum memories besides embeddability, bulk embeddability, and supermajorization. One natural example is embeddability in the presence of an auxiliary memory, or *stable* embeddability. Given memories $\mathscr{A}$ and $\mathscr{B}$, when is there a memory $\mathscr{C}$ such that

$$\mathscr{A} \otimes \mathscr{C} \hookrightarrow \mathscr{B} \otimes \mathscr{C}?$$

We do not know when $\mathscr{A}$ stably embeds in $\mathscr{B}$. Stable embeddability implies bulk embeddability and is implied by embeddability, but we do not know how it compares to supermajorization order.

Finally Theorem 1 is related to a much more general question in quantum information theory. Let $\mathscr{E} : \mathscr{A}_1 \to \mathscr{A}_2$ and $\mathscr{F} : \mathscr{B}_1 \to \mathscr{B}_2$ be quantum operations representing two quantum channels between general quantum memories. When are there operations $\mathscr{X}_N$ and $\mathscr{Y}_N$ that make the diagram

$$\begin{array}{ccc} \mathscr{A}_1^{\otimes N} & \xrightarrow{\mathscr{E}^{\otimes N}} & \mathscr{A}_2^{\otimes N} \\ \downarrow{\scriptstyle \mathscr{X}_N} & & \uparrow{\scriptstyle \mathscr{Y}_N} \\ \mathscr{B}_1^{\otimes N(1+\varepsilon)} & \xrightarrow{\mathscr{F}^{\otimes N(1+\varepsilon)}} & \mathscr{B}_2^{\otimes N(1+\varepsilon)} \end{array}$$

commute with high fidelity? We can then say that the channel $\mathscr{E}$ reliably bulk-encodes in the channel $\mathscr{F}$. Theorems 1, 2, and 7 together answer the question when $\mathscr{E}$ and $\mathscr{F}$ are both the identity map, with the refinement that perfect fidelity is possible when high fidelity is possible. In light of Theorem 4, it suffices to let $\mathscr{E}$ and $\mathscr{F}$ be CPU projections.

Acknowledgments

The author would like to thank Daniel Gottesman, Janko Gravner, Patrick Hayden, Dongseok Kim, and Bruno Nachtergaele for very helpful discussions.

[1] Howard Barnum, Patrick Hayden, Richard Jozsa, and Andreas Winter, *On the reversible extraction of classical information from a quantum source*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **457** (2001), no. 2012, 2019–2039, arXiv:quant-ph/0011072.

[2] Ola Bratteli, *Inductive limits of finite dimensional C^*-algebras*, Trans. Amer. Math. Soc. **171** (1972), 195–234.

[3] Carlton M. Caves, Christopher A. Fuchs, and Pranaw Rungta, *Entanglement of formation of an arbitrary state of two rebits*, Found. Phys. Lett. **14** (2001), no. 3, 199–212, arXiv:quant-ph/0009063.

[4] Man Duen Choi and Edward G. Effros, *Injectivity and operator spaces*, J. Funct. Anal. **24** (1977), no. 2, 156–209.

[5] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, 2nd ed., Springer-Verlag, New York, 1998.

[6] Richard V. Kadison and John R. Ringrose, *Fundamentals of the theory of operator algebras*, vol. I, Academic Press, 1983.

[7] ———, *Fundamentals of the theory of operator algebras*, vol. II, Academic Press, 1986.

[8] Michael A. Nielsen and Isaac L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge, 2000.

[9] Richard P. Stanley, *Enumerative combinatorics*, vol. 2, Cambridge University Press, 1999.

[10] Wojciech H. Zurek, *Environment-induced superselection rules*, Phys. Rev. D **26** (1982), no. 8, 1862–1880.